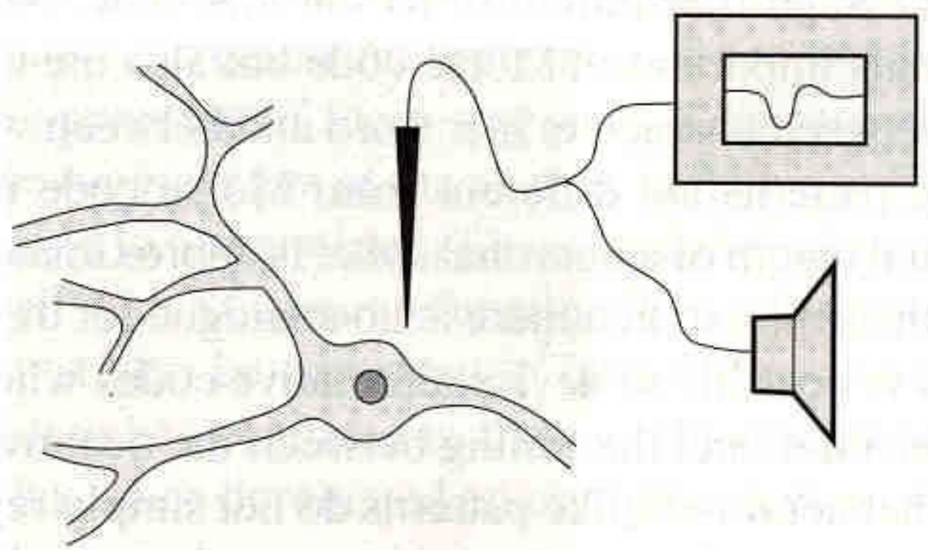
Teoría de información con aplicación a Representaciones y Código Neuronal

Introducción

- En este capítulo discutiremos cómo puede ser representada la información en el cerebro y como se puede analizar con herramientas basadas en Teoría de Información:
  - Spikes
  - Patrones de disparo neuronal
  - *Rate Coding*
  - *Spike Timing*, *Temporal Coding*
  - Teoría de Información

¿Cómo hablan las neuronas?

- Ya sabemos que podemos medir la actividad neuronal mediante registros de microelectrodos.



Schematic view of an extracellular recording of neuronal activity using a microelectrode.

¿Cómo hablan las neuronas?

- El comportamiento irregular en los disparos ha fascinado a científicos durante décadas.
- El disparo regular en principio no es interesante para el código neuronal (Predecible).
- La desviación de la regularidad transporta más información.
- La información en el tren de disparos sólo es interesante si es importante funcionalmente y si no es ruido.

¿Cómo hablan las neuronas?

- Con propósito ilustrativo, comparemos con el código morse que permitía transmitir un telegrama mediante impulsos eléctricos.

Table 5.1 International Morse code of short (.) and long (-) electrical signals, a variation of a code invented by Samuel Morse around 1835. The length of time for a 'dash' signal is set to be three times as long as a 'dot' signal. The time between signals within characters is equal to one dot signal, equivalent to three dot signals between characters, and the time gap between words is equal to seven dot signals.

A	.-	K	-.-	U	..-	4	....-
B	-...	L	.-..	V	...-	5	.....
C	-.-.	M	--	W	.--	6	-....
D	-..	N	-.	X	-..-	7	--...
E	.	O	---	Y	-.--	8	---..
F	..-.	P	.--.	Z	--..	9	----.
G	--.	Q	--.-	0	-----	Fullstop	.-.-.-
H	....	R	.-.	1	.----	Comma	--..--
I	..	S	...	2	..---	Query	..--..
J	.---	T	-	3	...--		

¿Cómo hablan las neuronas?

- En Morse no sólo es importante por la longitud de las señales individuales, sino también los tiempos entre señales que corresponden a diferentes caracteres.
- El código neuronal (CN) es completamente diferente, por ejemplo no hay analogía entre señal larga-corta, ya que la duración aproximada de los spikes es de un milisegundo y es más o menos constante debido a su forma estereotipada.
- El CN no representa un alfabeto como el morse.
- Es un código mucho más complicado.

¿Cómo hablan las neuronas?

La hipótesis del *Firing Rate*.

- Los Neurocientíficos intentan descifrar el CN buscando correlaciones fiables entre patrones de disparo y las consecuencias del comportamiento entre un estímulo y los patrones de actividad.
- Primer intento: E. D. Adrian (1920): el número de disparos subía con el aumento efectivo del estímulo.
- Esto corresponde a una subida en la tasa de disparo neuronal (*firing rate*).

¿Cómo hablan las neuronas?

La hipótesis del *Firing Rate*.

- Se puede estimar el promedio temporal de la tasa de disparos de una neurona con una ventana fija de tamaño ΔT,

$$\nu(t) = \frac{\text{número de disparos en } \Delta T}{\Delta T} =$$

$$= \frac{1}{\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} \delta(t' - t^f)\, dt'$$

- La variable t^f es el tiempo de disparo de la neurona.
- Así, para pequeñas ventanas se define la tasa de disparo instantánea.

¿Cómo hablan las neuronas?

La hipótesis del *Firing Rate*.

- La integral define la suma de los spikes dentro de la ventana de tamaño ΔT.
- La ventaja de esta formulación es que se puede usar para generalizar la tasa de disparo incluyendo pesos diferentes sobre la ventana temporal.
- Por ejemplo, es común usar una ventana gaussiana que no tiene fronteras abruptas como la ventana rectangular.

¿Cómo hablan las neuronas?

La hipótesis del *Firing Rate*.

- Así, la tasa de disparos a tiempo t mediante un ventana gaussina se define como:

$$\nu(t) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} \delta(t' - t^{f}) e^{(t'-t)^2/2\sigma^2} dt'$$

- Para estimar la tasa de disparo de un neurona, frecuentemente se realiza sobre varios experimentos con idénticas condiciones. Esto es debido a que el disparo de una neurona es altamente ruidoso.
- Esta medida es un procedimiento válido experimental para encontrar correlaciones con respuestas de comportamientos.

¿Cómo hablan las neuronas?

La hipótesis del *Firing Rate*.

- De esta manera, **Adrian** demostró que la **tasa de disparo** en un receptor elástico de la rana aumentaba según incrementaba el peso que se aplicaba al músculo.
- Generalmente las tasas de disparo neuronal aumentan significativamente en un tiempo relativamente corto cuando se presenta a la neurona un **estímulo efectivo**.
- El **campo receptivo** de una neurona es el conjunto de estímulos que, de una manera fiable, incrementan la frecuencia de disparos de la misma.

¿Cómo hablan las neuronas?

La hipótesis del *Firing Rate*.

- Otro ejemplo lo encontramos en la corteza visual primaria en la que la respuesta de una neurona responde a una orientación preferente de barras visuales (***tuning curves***).
- Estas curvas se pueden ajustar bastante bien por una función gaussiana.

¿Cómo hablan las neuronas?

La hipótesis del *Firing Rate*.

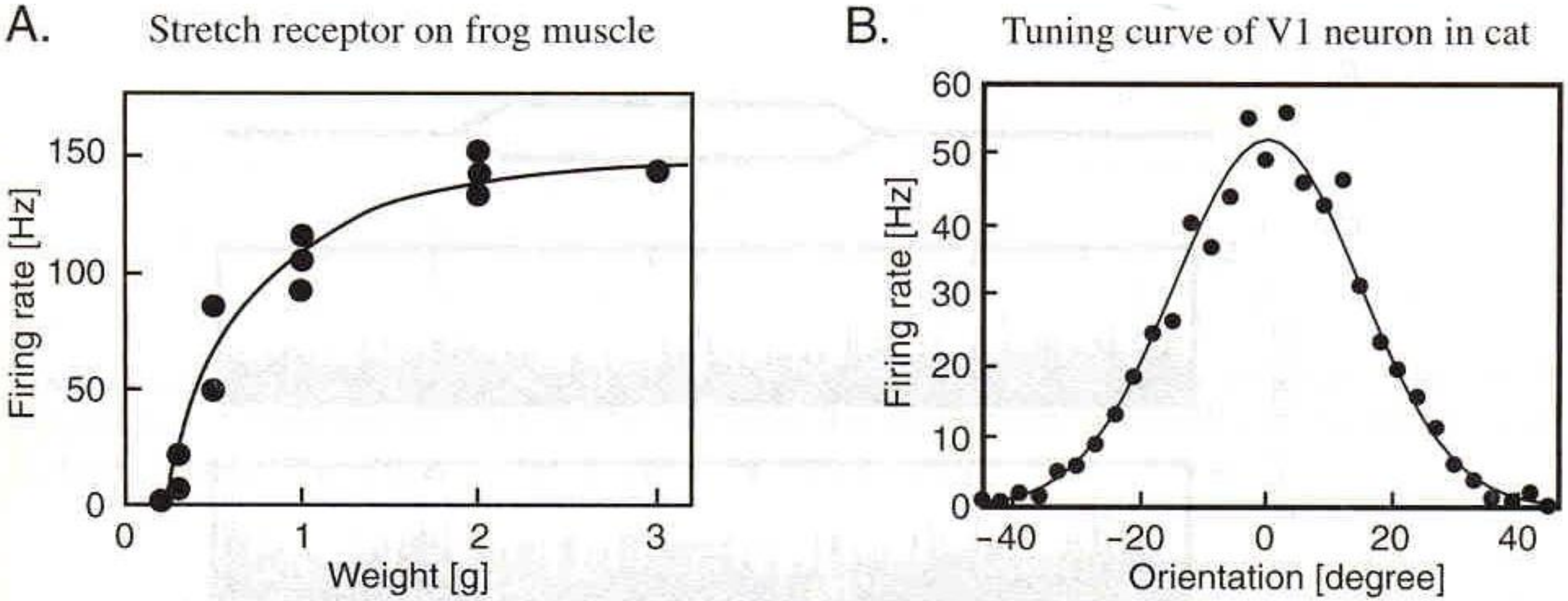


Fig. 5.2 (A) Data from Adrian's original work showing the firing rate of a frog's stretch receptor on a muscle as a function of the weight load applied to the muscle [redrawn from Adrian, *J. Physiol. (Lond.)* 61: 49–72 (1926)]. (B) Response (tuning curve) of a neuron in the primary visual cortex of the cat as a function of the orientation of a light stimulus in form of a moving bar [data from Henry, Dreher, and Bishop, *J. Neurophys.* 37: 1394–1409 (1974)].

¿Cómo hablan las neuronas?

Código de Correlación.

- El proceso de información en el cerebro normalmente incluye **variación de la frecuencia** de disparo.
- Así, esperamos que haya variación del número de *spikes* cuando variamos las corrientes de entrada.
- No obstante, queremos ver si existe algún tipo de **codificación diferente** a la tasa de disparo.

¿Cómo hablan las neuronas?

Código de Correlación.

- Un ejemplo peculiar es la respuesta en el **cortex auditivo primario** a un tono de 4kHz con una envolvente variable.
- La **frecuencia de disparo** de las dos neuronas **no muestran ningún cambio** significativo cuando variamos la amplitud de la señal de estímulo.
- Pero sí existe **correlación en frecuencia** de disparo relativa de una neurona a otra.

¿Cómo hablan las neuronas?
Código de Correlación.

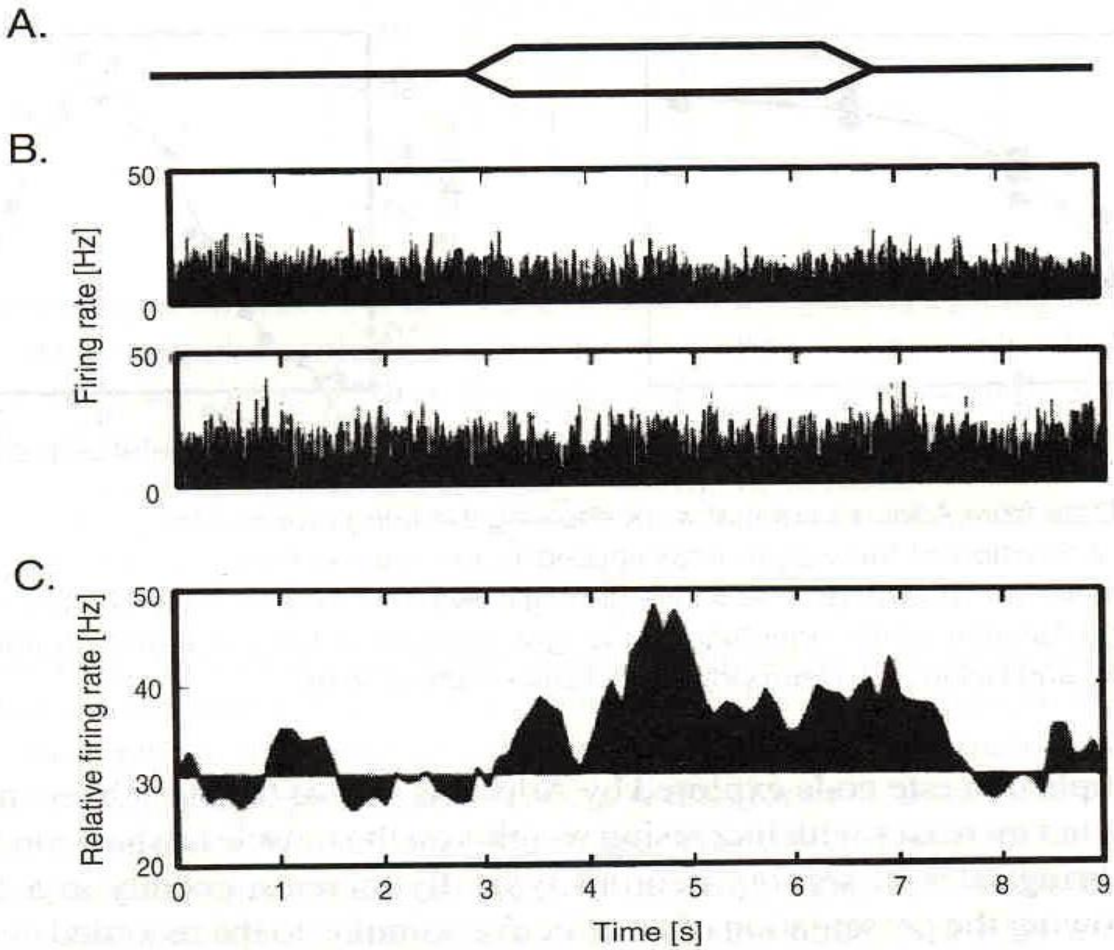


Fig. 5.3 An example of the response of some neurons in the primary auditory cortex that do not show significant variations in response to the onset of a 4 kHz tone with the amplitude envelope shown in (A). (B) Average firing rates in 5 ms bins of two different neurons. (C) Spike-triggered average rate that indicates some correlation between the firing of the two neurons that is significantly correlated to the presentation of the stimulus [from DeCharms and Merzenich, *Science* 381: 610–13 (1996)].

¿Cómo hablan las neuronas?

Integradores y Detectores de Coincidencia.

- En un **integrador perfecto** se puede generar un disparo por la acumulación de carga, así no se tiene mucho en cuenta la coincidencia de disparos.
- En contraste, en un **integrador con pérdida** (se puede usar como un detector de coincidencias cuando la constante de tiempo es pequeña) es fundamental el código de sincronización.
- Si el **umbral es mayor, se pueden detectar más de 2 disparos** (ver figura 5.4).

¿Cómo hablan las neuronas?

Integradores y Detectores de Coincidencia.

- En la siguiente figura ambas neuronas **disparan en el mismo instante** de tiempo.
- Pero el **motivo** por el que dispara cada una de ellas es **totalmente diferente**.
- En el primer caso, el disparo se debe a que han ocurrido cuatro disparos presinápticos. También podría haber disparado si uno de los dos disparos sincronizados hubiese ocurrido antes (ya que no hay perdida).
- En el segundo caso esto no ocurre, y el disparo se debe a la sincronización de los dos EPSPs.

¿Cómo hablan las neuronas?

Integradores y Detectores de Coincidencia.

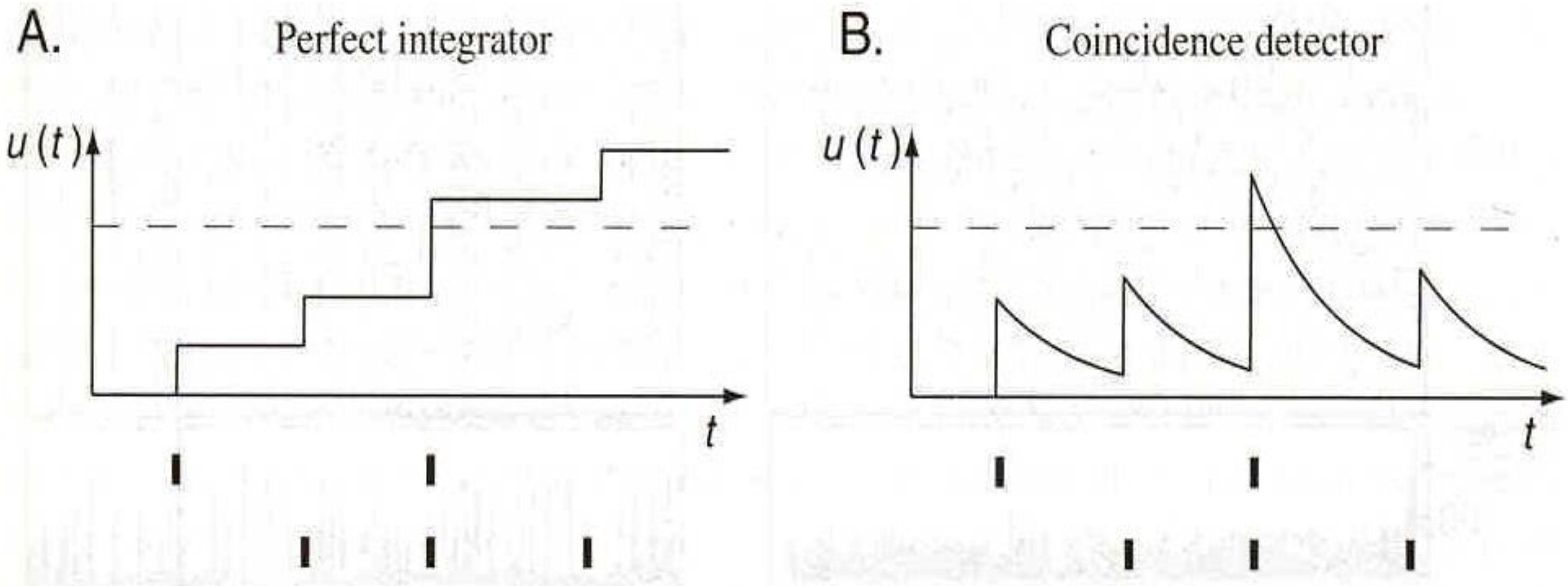


Fig. 5.4 Schematic illustration of (A) a perfect integrator and (B) a leaky integrator that can be utilized as coincidence detector. In this example the membrane potential $u(t)$ integrates short current pulses of the two spike trains shown at the bottom.

¿Cómo hablan las neuronas?

Precisión en la Coordinación de *spikes*.

- La **fiabilidad** o seguridad de transmisión de disparos es **baja** y se forma gran **variabilidad** en la respuesta neuronal.
- Un ejemplo es la respuesta de una neurona en el área temporal media (**MT**), que responde a los movimientos de un estímulo visual.
- Es más **preciso al principio**, pero luego empieza a ser más impreciso, lo que nos podría dar una idea de que el código neuronal es muy impreciso y variable.
- Sin embargo, este fenómeno lo tenemos que analizar en el contexto del experimento que estamos realizando.

¿Cómo hablan las neuronas?

Precisión en la Coordinación de *spikes*.

- Así, este fenómeno no ocurre cuando estimulamos con una señal variable rápidamente.
- Curiosamente, vemos que la **respuesta neuronal** sigue las **variaciones** en el **estímulo** con una escasa dispersión.
- No es que respondan todas la neuronas igual al cambio de pendiente, pero la mayoría de disparos siguen el cambio de pendiente.

¿Cómo hablan las neuronas?

Precisión en la Coordinación de *spikes*.

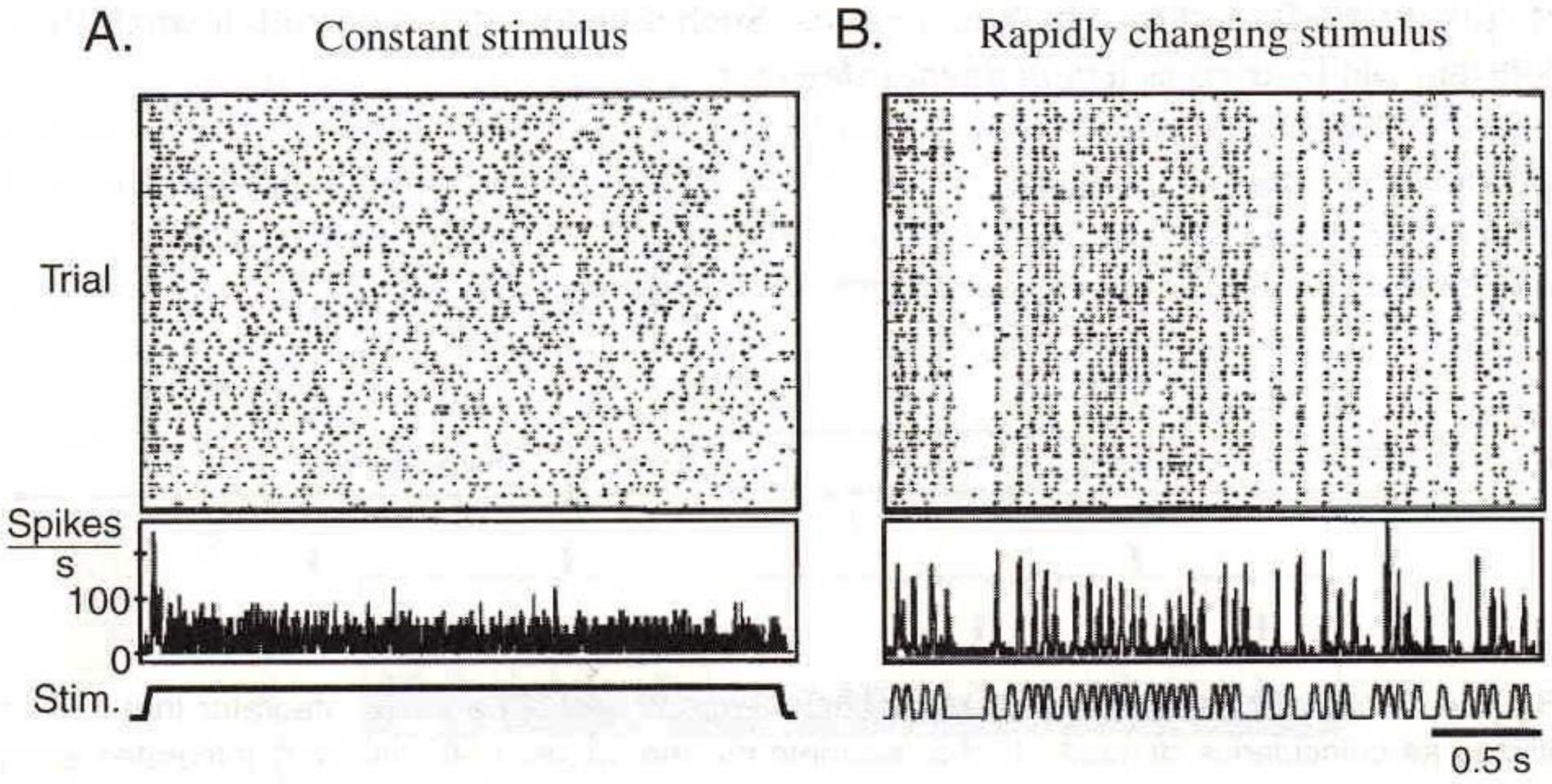


Fig. 5.5 Spike trains (top) and average response over trials (middle) of an MT neuron to a visual stimulus with either constant velocity (A) or altering velocity (B) as indicated in the bottom graph [adapted from Buračas *et al.*, *Neuron* 20: 959–69 (1998)].

¿Cómo hablan las neuronas?

Precisión en la Coordinación de *spikes*.

- Se puede estudiar el procesamiento de información en el cerebro apoyándonos en **teoría de información**.
- Podemos utilizar la teoría de información para comparar la información contenida en el patrón de disparos, con la información contenida en la frecuencia de disparos de los trenes de disparo.
- Veremos, aunque no siempre, que una gran cantidad de información es llevada por unos pocos potenciales de acción que hacen la **frecuencia de disparo altamente relevante**.

Teoría de Información:

Probabilidad y sus propiedades: Concepto

- Cuando tenemos una muestra de una población, después de describirlos, nuestro objetivo es inferir las propiedades de la población a partir de la muestra.
- El instrumento conceptual que permite la generalización es un modelo de la población.
- El calculo de probabilidades me permite calcula este modelo que actúa de puente entre lo observado, i.e. la **muestra**, y lo desconocido, i.e. la **población**.
- ¿Que es la probabilidad de una población finita homogénea de N elementos, k de los cuales tienen la característica A?
- $P(A)=k/N$

Teoría de Información:

Probabilidad y sus propiedades: Concepto

- Hay que ser consciente que muchas veces **no se puede obtener un conocimiento exacto** de la probabilidad ya que:
  - Al no ser posible una experimentación indefinida, siempre tenemos una información limitada sobre la frecuencia relativa.
  - El sistema observado puede variar a lo largo del tiempo, y por tanto también las frecuencias relativas.
  - Por tanto para poblaciones finitas la identificación de la probabilidad con la frecuencia relativa es casi inmediata, pero para poblaciones infinitas puede presentar serios problemas.
- Esto se complica mucho más para **sucesos** inciertos, que solamente **ocurren un número de veces muy reducido**.

Teoría de Información:

Probabilidad y sus propiedades: Definición

- Población: conjunto de elementos homogéneos en los que se desea investigar la ocurrencia de una característica o propiedad:
  - Numero de elemento finito o infinito.
  - Debe ser posible observar sus elementos.
  - Debe ser posible si un elemento pertenece a ella o no.
- Suceso elementales: es el conjunto de resultados posibles que verifican:
  - Siempre ocurre alguno de ellos.
  - Son mutuamente excluyentes.
  - Sucesos compuestos: los construidos a partir de uniones de resultados elementales.
  - P. ej. Tirar un dado:
  - Sucesos elementales: 1, 2, 3, 4, 5, 6
  - Sucesos compuestos: número par, número impar, menor 5, múltiplo de 2, etc.

Teoría de Información:

Probabilidad y sus propiedades: Definición

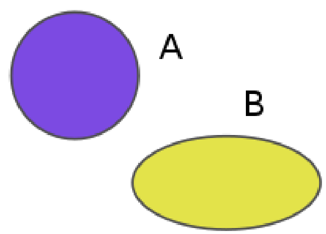
- Espacio muestral: conjunto de todos los resultados posibles del experimento. Así los elementos elementales y compuestos son los subconjuntos del espacio muestral.
- Definimos el suceso seguro, E, como todo el espacio muestral, es seguro porque siempre ocurre.
- Definimos el suceso imposible, Ø, como aquel que no ocurre nunca.

Teoría de Información:

Probabilidad y sus propiedades: Definición

- Se desea asociar a cada suceso una medida de incertidumbre que llamaremos probabilidad, con las siguientes propiedades:
- Debido a que la $f_r(A)$ es un numero comprendido entre 0 y 1, la probabilidad: $\boxed{0 \leq P(A) \leq 1.}$
- La frecuencia del suceso seguro ocurre siempre: $\boxed{P(E)=1.}$
- Si A y B son características mutuamente excluyentes y las unimos en una nueva C=A+B. C ocurre cuando ocurre A o ocurre B. La frecuencia relativa de C es la suma de las frecuencias relativas de A y B, y por tanto la probabilidad de sucesos excluyentes: $\boxed{P(A+B)=P(A \cup B)=P(A \text{ or } B)=P(A)+P(B).}$ P. ej probabilidad de sacar 1 o 2 en una tirada de dado: P(1 or 2) =1/6+1/6=1/3.

Imagen extraída de https://es.wikipedia.org/wiki/Probabilidad_condicionada



mutuamente excluyentes

Teoría de Información:

Probabilidad y sus propiedades: Definición

- Se desea asociar a cada suceso una medida de incertidumbre que llamaremos probabilidad, con las siguientes propiedades:
- Suponiendo A y B NO son mutuamente excluyentes definimos n_{AB}, $n_{A\overline{B}}$, $n_{\overline{A}B}$ al numero de veces que aparecen los sucesos mutuamente excluyentes (A y B), (A y no B), (no A y B), tendremos:
- $n_A = n_{AB} + n_{A\overline{B}}$, $n_B = n_{AB} + n_{\overline{A}B}$, $n_{A+B} = n_{AB} + n_{A\overline{B}} + n_{\overline{A}B}$, de estas 3 ecuaciones se puede obtener $n_{A+B} = n_A + n_B + n_{AB}$
- Si dividimos por el número total de observaciones obtenemos las frecuencias relativas que pasadas a probabilidades:
  - P(A+B)=P(A)+P(B)-P(AB)
  - P(A U B)=P(A)+P(B)-P(A∩B)
  - P(A or B)=P(A)+P(B)-P(A and B)
- P. ej. Probabilidad de elegir una carta al azar de una baraja sea un corazón o una cara (J, Q, K) es 13/52 + 12/52 – 3/52 =11/26 (de 52 13 son corazones, 12 caras , y 3 son las dos).

Teoría de Información:

Probabilidad y sus propiedades: Definición

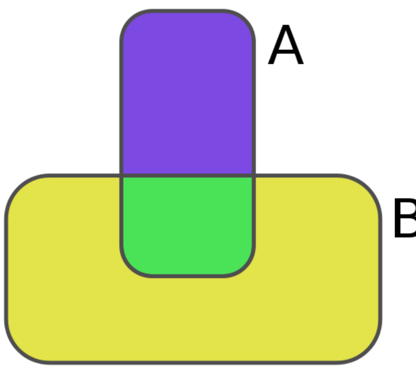
- Se desea asociar a cada suceso una medida de incertidumbre que llamaremos probabilidad, con las siguientes propiedades:
- Suponiendo que $\overline{A}$ es el suceso complementario de A, que ocurre siempre que no ocurre A, de las propiedades anteriores se puede deducir:
- $\boxed{P(\overline{A})=1-P(A).}$

Teoría de Información:

Probabilidad condicionada

- La frecuencia relativa de A condicionada a la ocurrencia de B se define considerando los casos en los que aparece B, y viendo en cuantos de estos caso ocurre A:
- $f_r(A|B)=n_{AB}/n_B$, como $f_r(A)=n_A/n$, $f_r(B)=n_B/n$ y $f_r(AB)=n_{AB}/n$, se tiene:
- $f_r(A|B)=f_r(AB)/f_r(B)$ o lo que es lo mismo:
- $=f_r(AB) = f_r(A|B)f_r(B)=f_r(B|A)f_r(A)$
- Así exigiremos la misma propiedad a la probabilidad y definiremos probabilidad de un suceso A condicionada a otro B como:
- $P(A|B)=P(AB)/P(B)$, ojo es una intersección en P(B)

Imagen extraída de https://es.wikipedia.org/wiki/Probabilidad_condicionada



Tomando los casos en los que B se cumple, la fracción en los que también se cumple A.

Teoría de Información:

Probabilidad condicionada, independencia

- Diremos que A y B son sucesos independientes, si el conocimiento de la ocurrencia de uno no modifica la probabilidad de aparición del otro:
- $P(A|B)=P(A)$ o $P(B|A)=P(B)$, o lo que es lo mismo:
- $P(AB)=P(A)\,P(B)$
- Esto se puede generalizar a cualquier numero de sucesos:
- Diremos que los sucesos $A_1,\ldots,A_n$, son independientes si la probabilidad conjunta de cualquier subconjunto que pueda formarse con ellos es el producto de las probabilidades individuales.

Teoría de Información:

Canal de Comunicación.

- Estamos hablando de Información, pero no hemos definido de una manera clara este concepto.
- Fue definido por primera vez por Claude **Shannon** en 1948, estudiando la transmisión de información en un canal ruidoso.

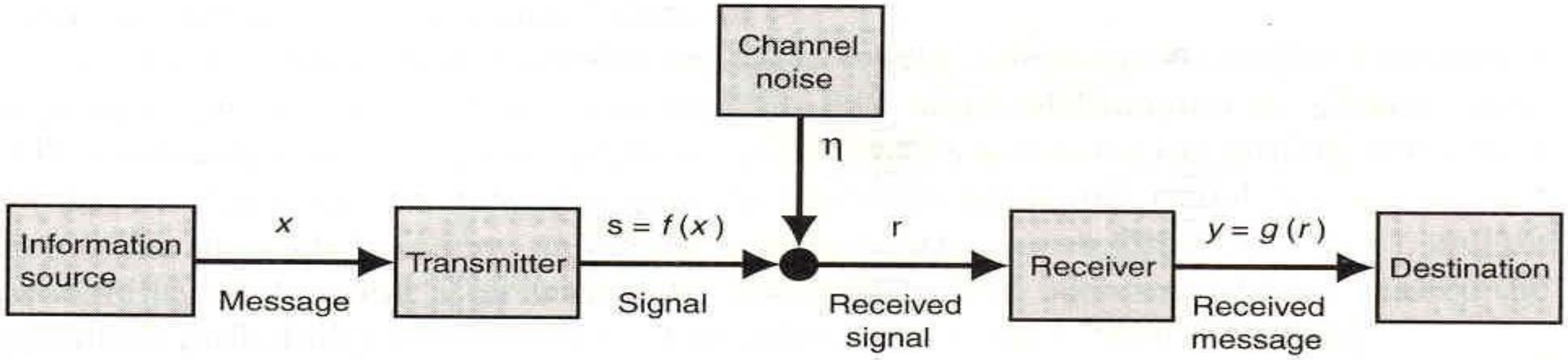


Fig. 5.6 The communication channel as studied by Shannon. A message x is converted into a signal $s = f(x)$ by a transmitter that sends this signal subsequently to the receiver. The receiver generally receives a distorted signal r consisting of the sent signal s convoluted by noise η. The received signal is then converted into a message $y = g(r)$ [adapted from C. Shannon, *The Bell System Technical Journal* 27: 379–423 (1948)].

Teoría de Información:

Canal de Comunicación.

- Un **mensaje x_i** de un conjunto de posibles mensajes X, se convierte en un señal $s_i=f(x_i)$.
- Se transmite sobre un **canal ruidoso**.
- El receptor recibe r_i y lo transforma a $\boldsymbol{y_i=g(r_i)}$ de un conjunto Y de posibles mensajes que puede interpretar el destinatario.
- En un canal **sin ruido $\boldsymbol{r_i=s_i}$ y $\boldsymbol{g=f^{-1}}$**.

Teoría de Información:

Ganancia de Información y Entropía.

- **¿Qué información ganamos cuando recibimos un mensaje y_i?**
- ¿Qué tiene más información?
  - Que os diga que me estoy comiendo una manzana de postre (más probable).
  - Que os diga que se está produciendo en estos momentos un terremoto en Cantoblanco (altamente improbable).
- Así, la información depende de los posibles mensajes y de sus frecuencias (probabilidad $p_i=P(y_i)$ del mensaje).

Teoría de Información:

Ganancia de Información y Entropía.

- Así, la información debería ser proporcional a la inversa de la probabilidad.
- Más aún, **diferentes informaciones independientes deberían ser aditivas**.
- Pero la probabilidad de mensajes independientes es el producto de las probabilidades.
- Por tanto, necesitamos una función que cumpla la propiedad: $\mathbf{f(xy)=f(x)+f(y)}$.
- El **logaritmo** tiene esas características.

Teoría de Información:

Ganancia de Información y Entropía.

- De esta manera, ya podemos definir nuestra medida de información como: $I(y_i)=\log_2(1/p_i)= -\log_2(p_i)$. Esto es la **ganancia de información** para un determinado mensaje.
- La **entropía** es el promedio de la información de todos los mensajes: $S(x)=-\sum_i p_i \log_2 (p_i)$.
- La entropía de un conjunto de N mensajes igual de probables es: $S(x)=-\sum_i (1/N) \log_2 (1/N) = =\log_2 (N)$.
- La ganancia de información se puede ver como una **reducción de la incertidumbre** del sistema: 9n+1b si sacamos 1b se elimina la incertidumbre.

Teoría de Información:

Ganancia de Información y Entropía.

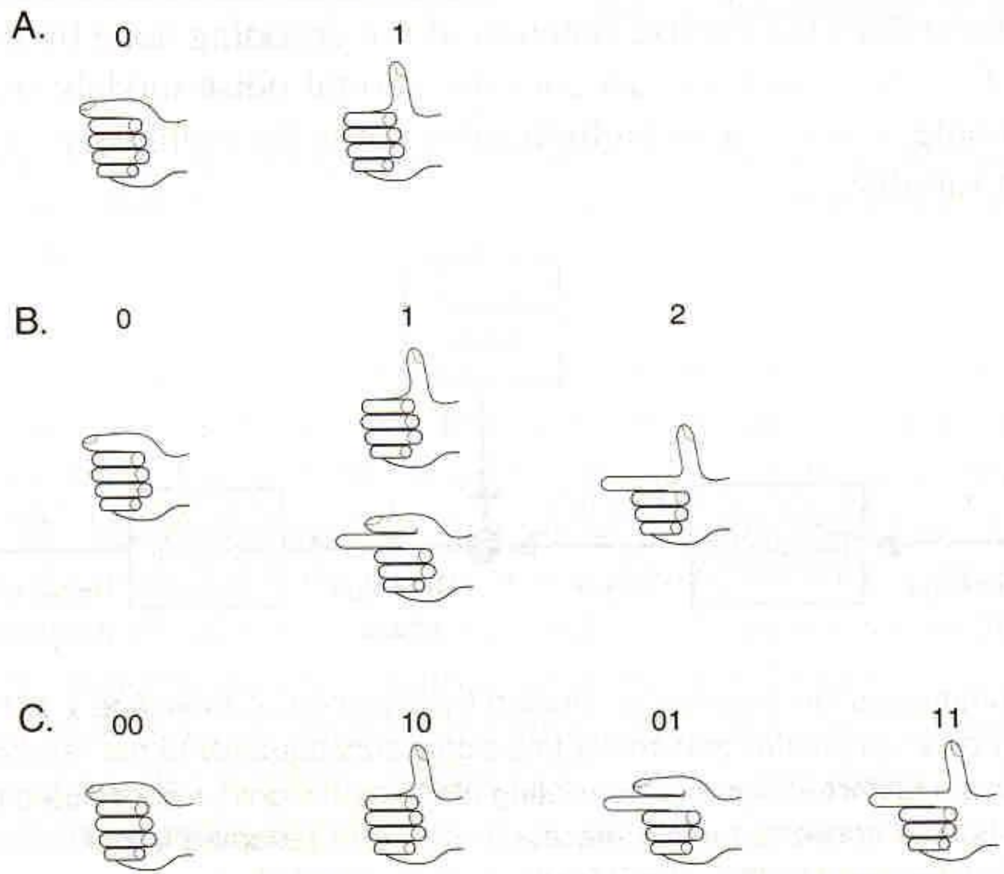


Fig. 5.7 Example of hand signals used to transmit messages. (A) Only two positions of the thumb are used to encode two possible messages '0' or '1'. (B) Two fingers are used to encode three messages by the total amount of extended fingers. (C) Two fingers are used with a different code, in which the position of each of the two fingers used for the signalling is essential.

Teoría de Información:

Ganancia de Información y Entropía.

- Así, en la figura anterior la ganancia de información para los diferentes casos, suponiendo las distribuciones de probabilidad equidistribuidas es:
  - Caso A: I= - $\log_2(0.5)$=1 bit. (0 o 1)
  - Caso B: I= - $\log_2(1/3)$=1.585 bits. (0, 1 o 2)
  - Caso C: I= - $\log_2(1/4)$=2 bits. (00, 01,10 o 11)

Teoría de Información:

Ganancia de Información y Entropía.

- Todo esto lo podemos extrapolar al campo continuo con distribuciones de probabilidad.
- Así, la entropía para una distribución gaussiana de media μ y varianza σ^2

$$p(x)=1/((2\pi)^{0.5}\,\sigma)e^{-(x-\mu)^2/(2\sigma^2)}$$

$$S=0.5\log_2(2\pi e\sigma^2)$$

(demostrar esto como ejercicio para casa (1))

- La entropía depende de la varianza y no de la media de la gaussiana.

Teoría de Información:

Información Mutua.

- Hasta el momento hemos considerado transmisión de información libre de ruido ($g=f^{-1}$ en la última figura).
- Pero en sistemas reales esto no pasa, y nos tenemos que preguntar qué nos puede decir la señal “y” sobre la señal que se quería enviar “x”.
- Para eso definimos una nueva cantidad denominada información mutua.
- Esta nueva cantidad mide la reducción de la incertidumbre (entropía) de una variable aleatoria, X, debido al conocimiento del valor de otra variable aleatoria Y.
- Es decir mide la dependencia mutua de las dos variables aleatorias X e Y.

Teoría de Información:

Información Mutua.

- Así definimos para dos variables aleatorias $X=\{x_i\}$ y $Y=\{y_i\}$, con i=1…n, la información mutua entre el suceso x_i e y_i como

$$I(x_i;\, y_i)= \log_2 (p(x_i \mid y_i)/(p(x_i))$$

- Así la información mutua media o promedio la podemos ver como

$$MI(x,y)= \sum_{xy} p(x,y)\, I(x;\, y)$$

$$MI(x,y)= \sum_{xy} p(y)p(x|y) \log_2 (p(x|y)/(p(x))\,)=$$

$$\sum_{xy} p(x,y) \log_2 (p(x,y)/(p(x)p(y))\,)$$

(ya que $P(x|y)=P(x,y)/P(y)$)

Demostrar para casa que $MI(x,y)=MI(y,x)$ (2)